ERIC Notebook

May/June 2003 Issue 28

UNC SCHOOL OF PUBLIC HEALTH
DEPARTMENT OF EPIDEMIOLOGY

Lorraine Alexander, DrPH
Director of Education Program

Katherine Hartmann, MD, PhD

Victor Schoenbach, PhD

Sandra Deming, MSPH
Felisha Griffin, MSPH
George Jackson, MHA
Richard MacLehose, MS
Graduate Research Asst.

http://www.sph.unc.edu/courses/eric

VA ERIC
EPIDEMIOLOGIC RESEARCH & INFORMATION CENTER
Durham, North Carolina

Eugene Oddone, MD, MHSc
Acting ERIC Director

Beth Armstrong
ERIC Program Manager

http://hsrd.durham.med.va.gov/ERIC/

The ERIC Notebook is funded by the Department of Veterans Affairs (DVA), Office of Research and Development (ORD), Cooperative Studies Program (CSP), to promote the strategic growth of the epidemiologic capacity of the DVA.

Health Care Epidemiology: Meta-Analysis

A prior *ERIC Notebook* (issue 26) discussed evidence-based medicine, including the process of systematic reviews. This issue outlines the quantitative and analytic extension of systematic reviews: meta-analysis. We provide information about software and text that will allow readers to perform a meta-analysis.

Traditionally, when scientists have attempted to reach conclusions from a number of studies related to a given research question, they have used informal qualitative techniques that fail to systematically combine the results of all of the relevant studies. With the development of meta-analyis, we have a quantitive technique to combine the results of individual studies to obtain an overall estimate of effect size of a given intervention or exposure to evaluate the quality of a body of research literature.[1]

Often, meta-analysis allows researchers and clinicians to extract information that would otherwise not be apparent from the results of individual studies. This means that sometimes the meta-analysis will reveal the existence of an effect where individual studies lack the statistical power to do so or it may mean that the promise implied by an individual study is not realized when the results of multiple studies are quantitatively combined. Properly used meta-analyis allows for firmer quantitive grounds on which to base clinical guidelines, administrative, actions, or policy decisions.

Steps in Conducting a Systematic Review and Meta-Analysis

1. Determine the Research Question

Just as with any research study, the first step of a meta-analysis should be to formulate a well-defined research question. Are you interested in a summary estimate of "What dose of a drug is most efficacious?" or are you interested in identifying the sources of heterogeneity among studies that may contribute to differing results? An example of this type of question is "What characteristics of these clinical trials of a specific hypertension medication have led to the different outcomes?"

2. Identify Relevant Studies

Meta-analysis requires the determination of all potentially relevant studies. Failure to do so can lead to biases and inaccurate results.

To ensure that a literature search is as complete as possible, multiple search strategies should be employed. Start by consulting a professional. Librarians can point out multiple and perhaps unfamiliar sources of results. They may also provide search strategies that can enhance the completeness of your literature search using sources such as MEDLINE, the Cochrane Collaboration Clinical Trials Register[2], and other electronic databases. Additional published works can be selected through the bibliographies of previously identified publications.

It is also important to include studies that may not have been published due to null or unexpected results[3]. This "publication bias[3, 4]" can lead to inaccuracies in the estimation of associations due to missing data in your meta-analysis.[5] Talking to experts in the field can also aid in finding studies that are more recent or yet unpublished.

http://www.sph.unc.edu/courses/eric

http://www.sph.unc.edu/courses/eric

The **ERIC Education Website** has a number of free online epidemiology education resources. These include:

- Free access to all *ERIC Notebook* issues
- Form for free subscription to *ERIC Notebook*
- Education certificates based on issues of *ERIC Notebook*
- Free epidemiology education modules with certificates of completion
- Free epidemiology case studies
- Links to other online epidemiology education resources

3. Determine Inclusion and Exclusion Criteria and Select Appropriate Studies.

Inclusion and exclusion criteria for the studies should be determined based upon the research question being asked in the meta-analysis. A meta-analysis may have stringent inclusion and exclusion criteria if you want a specific answer in a very defined population. Furthermore, one should consider such factors as study design, methodology, and populations when defining criteria.[6] In instances where one is interested in identifying factors leading to the conflicting results among studies, the inclusion and exclusion criteria should be limited to ensure the inclusion of as many relevant studies as possible so that the sources of heterogeneity can be explored.

Lastly, it is important to exclude those studies, which may be duplicate reports to avoid counting the same estimates twice. It is not unusual to find the same study population evaluated for various factors with some overlap of data and results.

4. Assess Quality of the Individual Studies

When selecting studies to be included in the meta-analysis, it is necessary to assess the quality of individual articles. Lohr and Carey define methodologic quality as "the extent to which all aspects of a study's design and conduct can be shown to protect against systematic bias, nonsystematic bias, and inferential error" (p. 421).[7] In other words, the quality of your meta-analysis may be reduced by the poor quality of component studies. Thus, a method needs to be employed to ensure inclusion of all relevant studies of adequate quality.

Selecting scales or checklists to use in quality assessment should be done with great care. Decisions to include or weight studies based on quality in a meta-analyis can be quite different depending on the quality rating system that is selected. Different quality scales can even cause meta-anlyses on the same topic to reach opposite conclusions.[8]

A 2002 systematic review of over 100 mechanisms for assessing the quality of research studies conducted for the Agency for Healthcare Research and Quality (AHRQ) suggests that when selecting a system to rate quality, investigators should consider the type of studies being reviewed, important methodologic issues related to the topic of the meta-analysis, time needed to utilize the systems, and whether there is a preference for using a quality scale or checklist. The authors also determined the important domains that should be included in scales or checklists and identified specific scales and checklists that cover key domains. This information can be accessed on the AHRQ World Wide Web site (http://www.ahrq.gov/clinic/tp/strengthtp.htm).[9]

5. Extract Data

Once studies have been identified for inclusion in the meta-analysis, data related to the question of interest should be extracted from the studies. Ideally, data extraction should be carried out by more than one observer, a content and methodologic expert, which allows for a more nuanced appraisal of the literature. When practical, it is recommended that the observers be blinded to the journals, authors, and their supporting institutions to allow for data extraction without interpretation. In instances where there are discrepancies in the extracted data, a third observer may be employed to reduce the inclusion of erroneous or biased data.

In a perfect setting, results from all studies would be comparable because their estimates would be in similar forms, (e.g., odds ratios or risk differences). However, this is rarely the case. Even in instances where the same types of estimates are presented, the results may be from regression analyses that have not adjusted for the same factors. The inconsistencies in data and results presentation may require the meta-analyst to back-calculate to obtain "raw" data that is in a form suitable for aggregation. In cases where such calculations are not possible due to how the results are presented, personal contact with the author of the study is advisable to obtain suitable data. Every reasonable effort should be made to collect the necessary data, and thus to be as inclusive as possible which will allow for the evaluation of homogeneity among the studies.

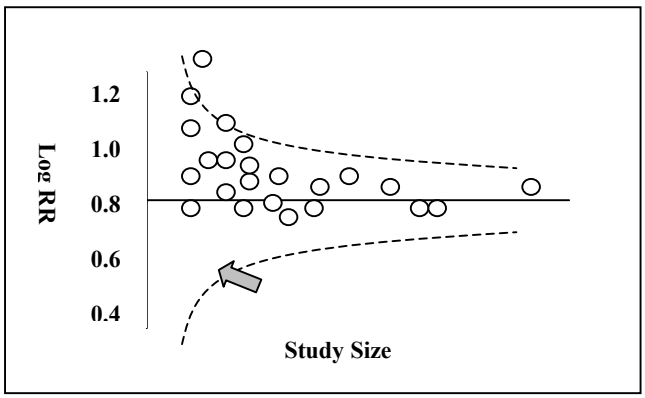
6. A Method for Assessing the Completeness of Literature Search

Once a researcher has consulted the experts and concluded the literature search, how do you determine how complete the search is? A graphical evaluation called a funnel plot can address this issue.[3, 10] A funnel plot evaluates whether there is a lack of publication of certain types of studies, namely those with null or inconsistent results. Below is an example of a funnel plot, which graphs the studies by sample size and result (log odds ratio, log relative risk, etc). The y-axis might also be the log odds ratio (log OR) or the standard error of the log OR. Let us assume that these are clinical trials evaluating the risk of neutropenia following a new chemotherapy thought to reduce the incidence of this side effect. The centerline represents the combined estimate. In the absence of publication bias, the dotted lines represent the expected funnel shape where larger studies have less variation around the overall estimate, and smaller studies have greater variation around the overall estimate.

The funnel plot below suggests that there may be some publication bias in relation to studies of interest in our meta-analysis. The lower left-hand portion of the graph (see arrow) seems to lack small studies that

demonstrate no effect of this drug, or an increased incidence of neutropenia. These types of results might be

Example Funnel Plot



RR=Relative Risk

expected to be more difficult to publish. In this instance, the meta-analyst should determine if there are unpublished studies of this nature that may have been missed in the initial literature search. Even if you don't identify additional studies, this possibility should be noted.

7. Analyze Data

Once researchers have extracted data that can be combined to calculate an overall mean, odds ratio, or risk ratio, there is a tendency to produce such a value regardless of whether it is appropriate. As indicated earlier, there is often a great deal of heterogeneity among studies that can make the produced overall estimate meaningless or, at the very least, uninterpretable. It is not very difficult to find examples in the meta-analysis literature where evaluation of heterogeneity indicated that it was inappropriate to combine the data, yet an overall estimate was produced and interpreted. This is especially concerning when issues such as developing clinical guidelines or policy are at stake. Furthermore, producing a combined estimate may lead to a failure to detect important clinical differences that may influence the individual results, further underscoring the importance of considering heterogeneity among study results.[11]

How does one determine which data can be appropriately combined?[12-14] A test of homogeneity can provide this information by assessing whether individual study results are likely to reflect a single underlying effect or a distribution of effects. If the test fails to detect heterogeneity (reject the null hypothesis that the estimates in the studies are homogeneous), then we are assuming that the differences between the studies are due to chance and sampling variation.

A chi-square test is commonly used to test the homogeneity of the individual study results.[13, 15] A test of homogeneity will produce a *Q* statistic that has a chi-square distribution with *k*-1 degrees of freedom. The *Q* statistic has an associated p-value, which is utilized to determine if there are significant sources of heterogeneity among the studies. Generally, a p-value equal or less than 0.05 is the cut-point under which the null hypothesis can be rejected. However, a p-value greater than 0.05 should not be seen as evidence for homogeneity since tests of homogeneity typically have low power and may fail to statistically detect a moderate amount of true heterogeneity among the studies. Often tests of homogeneity have a higher threshold for rejection ($p \leq 0.10$ or $p \leq 0.20$) to account for the low power due to the generally small sample size (number of individual studies).[15] Furthermore, one should remember that the 0.05 or 0.20 p-value is an arbitrary cut-point, and should never serve as a replacement for common sense and substantive knowledge of the subject under investigation. A positive test for heterogeneity should, however serve as the impetus for evaluating the factors that may have led to the clinical differences among the studies.

A more formal examination of heterogeneity can be accomplished using meta-regression techniques, which involve an application of linear regression to meta-analysis. This allows the meta-analyist to have a much more exact idea of the relative causes of study differences or similarities. Fortunately, software exists which allows for these calculations. The effect estimates (e.g. ln(OR), ln(RR), or RD) are treated as dependant variables. Including study characteristics as independent variables in the meta-regression and examining the coefficient for each study characteristic is a straightforward way to explore heterogeneity.[16]

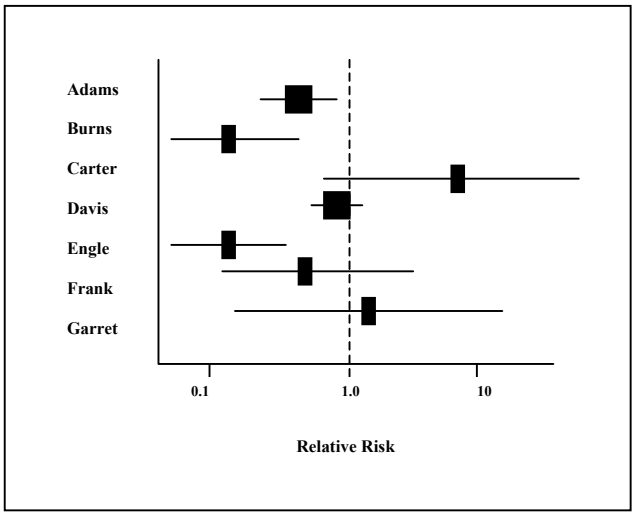
For example, if one were analyzing a systematic review where some studies were case-control and some were cohort and where some provided odds ratios adjusted for gender and some did not, the analyst could use meta-regression to decide whether study design or adjustment for gender caused a change of the ln(OR). As in linear regression, a one-unit change in the study characteristic is accompanied by a coefficient-unit change in the dependant variable, the effect measure. Meta-regression software also often calculates the amount of residual between-study variation, after the independent variables are controlled for, so the analyst can determine whether there is still an appreciable amount of heterogeneity.

Evaluating the differences among the studies can be done via two methods.[14] The first is tabular form. It is customary for Table 1 in meta-analyses to list the individual studies, the data/results of interest (such as an odds ratio), and the factors that may be important to the differences in the results (blinding, study design, stage of disease, etc.). Graphical displays of the data from the individual studies are another means of evaluating the similarities or differences of studies.[10, 17, 18] Below is one example of a graphical display: the forest plot.[10]

The individual studies are presented on the y-axis with the relative risk presented on the x-axis. The size of the box is proportional to the individual study sample sizes, while the horizontal lines typically correspond to the 95%

confidence intervals for the relative risk estimates. Sometimes the studies may be grouped on the y-axis, particularly when there are meaningful differences among

Sample Forest Plot



the studies that may be responsible for the heterogeneity in the results. Assume that the forest plot above represents the relative risk of neutropenia in response to a new chemotherapeutic agent as compared to a current regimen. For instance, the fictitious studies by Carter and Garret in the above forest plot demonstrate relative risks greater than one, while the other studies suggest a reduction in the incidence of neutropenia. A meta-analyst should look for reasons behind the differences between these groups. Perhaps the Carter and Garret studies selected study populations with different types of cancer or with more late stage disease as compared to the others. Furthermore, we can see how the evaluation of the differences may provide a more meaningful and interpretable result when determining for whom this drug is most effective in reducing the incidence of neutropenia.

However, one may still wish to produce a combined estimate of effect sizes for the entire group or for certain subgroups. Specific methods of calculating combined effect sizes are beyond the scope of this *ERIC Notebook*. The text *Systematic Reviews in Health Care*[19] and the publication by Breslow and Day[13] are good starting points for more general methods and formulae.

In general, a summary statistic is produced from measures of effect or association calculated for each individual study. A weighted average is then calculated for those summary estimates with the individual study weight typically being the inverse of the study variance.[20] The variance is closely related to sample size. Thus, larger studies will be given more weight in the overall estimate. The methods of weighting can change depending upon certain circumstances. Some examples include having small sample sizes, an imbalance in the allocation of study subjects, or rare events.

As alluded to earlier in this Notebook, calculating a summary statistic may be no longer sufficient in a meta-analysis. A complete meta-analysis should include a sensitivity analysis.[21] A sensitivity analysis can involve repeating the analysis on subsets of the original data as well as determining how any one study might influence the overall summary statistics. One very large study might have a profound, and perhaps misleading influence on the overall result. Sensitivity analysis may provide insight into the individual study factors that can affect the results, and that may be important to consider in future studies. This is the point at which personal experience and substantive knowledge can help in the analyses and interpretation.

For those interested in performing their own meta-analyses, statistical software developers now have packages that can produce many of the aforementioned calculations and graphical presentations. The Stata™ statistical package offers comprehensive tools for meta-analyses and includes functions that can generate summary statistics, funnel plots, and sensitivity analyses.[22] The SAS™ Institute also provides a commonly utilized statistical package for conducting meta-analysis.[23] Macros specifically related to meta-analysis can be accessed via the SAS™ Website (http://www.sas.com). A listing of additional meta-analysis software can be found on the *BMJ* Website referenced at the end of this Notebook.

8. Interpret Results

Now that all of the calculations have been preformed, the meta-analyst needs to fully discuss the limitations of the analysis, including potential biases and reasoning behind presenting or not presenting summary estimates. The factors that have led to any heterogeneity among the studies should be also presented.[11, 19, 24] It is important to indicate which studies may have been excluded from the analyses and the reasons for such a decision.

Furthermore, it is critical to reflect on the original study question.[19] Was the purpose of this analysis to generate a summary statistic that may have clinical practice or policy implications or was it to investigate the factors that may have led to previously inconsistent results, so that they may be addressed in future studies? The implications of the research question should be clearly discussed so that the reader can interpret the results within the appropriate context.[25]

Conclusion

Meta-analyis provides a method for combining study results to identify trends in those results and determine sources of inconsistences. The goal is to draw better conclusions about the meaning and coherence of a body of literature on a given topic. This Notebook has introduced some of the principal concepts involved in meta-analysis. Extended summaries of the topic include Stamps (2002)[1] and Greenland, (1998).[16] A number of

comprehensive books, including Egger, et al. (2001)[19] and Cook, et al. (1994)[26], may be useful.

Helpful Web Sites

AHRQ Evidence Report- *Systems to Rate the Strength of Scientific Evidence, Summary*
http://www.ahrq.gov/clinic/tp/strengthtp.htm

BMJ [British Medical Journal] Collection on Systematic Reviews and Meta- Analysis-Descriptions (must have access to BMJ Web site)
http://bmj.com/cgi/collection/systematic_reviews%3Astatistics_descriptions

BMJ [British Medical Journal] Collection on Systematic Reviews and Meta- Analysis-Examples (must have access to BMJ Web site)
http://bmj.com/cgi/collection/systematic_reviews%3Astatistics_examples

Campbell Collaboration
http://www.campbellcollaboration.org

Cochrane Collaboration
http://www.cochrane.org

MEDLINE
http://www.pubmed.gov

Research Triangle Institute-University of North Carolina Evidence-based Practice Center
http://www.rti.org/epc

SAS Institute
http://www.sas.com

Stata Statistical Software
http://www.stata.com

References

1. Stamps, A. E. (2002). Meta-analysis. In R. Bechtel and A. Churchman, (Eds.). *Handbook of Environmental Psychology* (pp. 222-232). New York: Wiley.

2. The Cochrane Controlled Trials Registrar. In: *The Cochrane Library*. *Oxford: Update Software*.

3. Thornton, A. & Lee, P. (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology, 53*, 207-216.

4. Easterbrook, P. J., Berlin, J. A., Gropalan, R., & Matthews, D. R. (1991). Publication bias in clinical research. *Lancet,337*, 867-872.

5. Sterling, T., Rosenbaum, W., & Weinkam, J. (1995). Publication decisions revisited: The effect of the outcome of statistical test of significance - or vice versa. *Journal of the American Statistical Association, 54*, 30-34.

6. Counsell, C. (1997). Formulating questions and locating primary studies for inclusion in systematic reviews. *Annals of Internal Medicine, 127*, 380-387.

7. Lohr, K. N. & Carey, T. S. (1999). Assessing "best evidence": Issues in grading the quality of studies for systematic reviews. *Joint Commission Journal on Quality Improvement, 25*, 470-479.

8. Juni, P., Witschi, A., Bloch, R., & Egger, M. (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA, 282*, 1054-1060.

9. Research Triangle Institute-University of North Carolina Evidence-based Practice Center. (2002,March). *Systems to Rate the Strength of Scientific Evidence, Summary (Evidence Report/Technology Assessment Number 47)* (AHRQ Publication No. 02-E015). Rockville, MD: AHRQ Publications Clearinghouse.
[http://www.ahrq.gov/clinic/tp/strengthtp.htm].

10. Light, R. J. & Pillemer, D. B. (1984). *Summing Up. The science of reviewing research*. Cambridge: Harvard University Press.

11. Thomspon, S. G. (1994). Why sources of heterogeneity in meta-analysis should be investigated. *British Medical Journal, 309*, 1351-1355.

12. Deeks, J. & Altman, D. (2001). Effect measures for meta-analysis of trials with binary outcomes. In: M. Egger, G. D. Smith, & D. G. Altman, (Eds.). *Systematic Reviews in Health Care: Meta-Analysis in Context* (2nd ed., pp. 313-335). London: BMJ Publishing Group.

13. Breslow, N. E. & Day, N. E. (1980). Combination of results from a series of 2 X 2 tables: control of confounding. *Statistical methods in Cancer Research: The Analysis of Case-Control Data. IARC Scientific Publications No. 32.* (Vol. 1). Lyon: International Agency for Health Research on Cancer.

14. Deeks, J., & Altman, D., & Bradburn, M. (2001). Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: M. Egger, G. D. Smith, & D. G. Altman, (Eds.). *Systematic Reviews in Health Care: Meta-Analysis in Context* (2nd ed., pp. 313-335). London: BMJ Publishing Group.

15. Petitti, D. B. (2000). *Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis, Methods for Quantitative Synthesis in Medicine* (2nd ed.). New York: Oxford University Press.

16. Greenland, S. (1998). Meta-analysis. In K. J. Rothman & S. Greenland (Eds.), *Modern Epidemiology* (2nd ed., pp. 643-673).Philadelphia, PA: Lippincott-Raven.

17. Galbraith, R. F. (1987). A note on graphical presentation of estimated odds ratios from several clinical trials. *Statistics in Medicine,7*, 889-894.

18. L'Abbe, K. A., Detsky, A. S., & O'Rourke, K. (1987). Meta-analysis in clinical research. *Annals of Internal Medicine, 107*, 224-233.

19. Egger, M., Smith, G. D., & Altman, D. G. (Eds.). *Systematic Reviews in Health Care: Meta-Analysis in Context* (2nd ed.). London: BMJ Publishing Group.

20. Hedges, L. V. & Olkin, I. (1985). *StatisticalMethods for Meta-Analysis*. Orlando: Academic Press.

21. Hulley, S, B., Cummings, S. R., Browner, W. S., Grady, D., Hearst, N., & Newman T. B. (2001). *Designing Clinical Research: An Epidemiologic Approach* (2nd ed.). Philadelphia: Lippincott Williams & Wilkins.

22. STATA. College Station, TX: Stata Corporation.

23. SAS. Cary, NC: SAS Institute.

24. Oxman, A. D. & Guyatt, G. H. (1992). A consumer's guide to subgroup analyses. *Annals of Internal Medicine, 116*, 78-84.

25. Gelber, R. D. & Goldhirsch, A. (1993). From the overview to the patient: how to interpret meta-analysis data. *Recent Results in Cancer Research, 127*, 167-176.

26. Cook, T. D., Cooper, H., Cordray, D. S., Hartman, H., Hedges, L. V., Light, R. J., Louis, T. A. & Mosteller, F. (Eds.). (1994). *Meta-analysis for Explanation: A* . New York: Russell Sage.

ERIC NOTEBOOK IS PRODUCED BY THE EDUCATIONAL ARM (LORRAINE K. ALEXANDER, DRPH, DIRECTOR) OF THE EPIDEMIOLOGIC RESEARCH AND INFORMATION CENTER AT DURHAM, NC (EUGENE Z. ODDONE, MD, MHSC, ACTING DIRECTOR).

If you are not currently receiving ERIC Notebook and would like to be added to the distribution list, please fill out the form below:

Name: _____

Degree(s):_____

Address:_____

City, State, Zip:_____

Telephone Number:_____

Fax Number:_____

E-mail Address:_____
Affiliation: VA_____ Other Gov't_____
Academic_____ Private_____

Please fax to: 919-416-5839 – Attn: Beth Armstrong
Or Email: beth.armstrong@duke.edu
Mail to: Beth Armstrong, ERIC Program Manager, VA Medical Center (152), 508 Fulton Street, Durham, NC 27705

Upcoming Topics

- **Disparities in Health Care**
- **Rapid Cycle Improvement**
- **Diabetes and Disparities**

Please let Beth Armstrong know which topics are of special interest to you so that we can include them in future issues:

beth.armstrong@duke.edu

Reminder:
Online **education certificates** are available for each ERIC Notebook (1 hour of AMA Category 2 CME credit). ERIC notebooks and education certificate questions can be viewed online at:

http://www.sph.unc.edu/courses/eric

ERIC notebooks can also be viewed online at:

http://hsrd.durham.med.va.gov/ERIC/Education/Education.html

BETH ARMSTRONG, ERIC PROGRAM MANAGER
VA MEDICAL CENTER (152)
508 FULTON STREET
Durham, NC 27705